

Schladitz, Sandra; Groß Ophoff, Jana; Wirtz, Markus

Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz

Blömeke, Sigrid [Hrsg.]; Zlatkin-Troitschanskaia, Olga [Hrsg.]: *Kompetenzen von Studierenden*. Weinheim u.a. : Beltz Juventa 2015, S. 167-184. - (Zeitschrift für Pädagogik, Beiheft; 61)



Quellenangabe/ Reference:

Schladitz, Sandra; Groß Ophoff, Jana; Wirtz, Markus: Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz - In: Blömeke, Sigrid [Hrsg.]; Zlatkin-Troitschanskaia, Olga [Hrsg.]: *Kompetenzen von Studierenden*. Weinheim u.a. : Beltz Juventa 2015, S. 167-184 - URN: urn:nbn:de:0111-pedocs-155098 - DOI: 10.25656/01:15509

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-155098>

<https://doi.org/10.25656/01:15509>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

61. Beiheft

April 2015

ZEITSCHRIFT FÜR PÄDAGOGIK

**Kompetenzen
von Studierenden**

BELTZ VERLAG **JUVENTA**

Zeitschrift für Pädagogik · 61. Beiheft

Kompetenzen von Studierenden

Herausgegeben von

Sigrid Blömeke und Olga Zlatkin-Troitschanskaia

BELTZ JUVENTA

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, bleiben dem Beltz-Verlag vorbehalten.

Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2015 Beltz Juventa · Weinheim und Basel

www.beltz.de · www.juventa.de

Herstellung: Lore Amann

Satz: text plus form, Dresden

E-Book

ISSN 0514-2717

Bestell-Nr. 443508

Inhaltsverzeichnis

<i>Sigrid Blömeke/Olga Zlatkin-Troitschanskaia</i> Kompetenzen von Studierenden. Einleitung zum Beiheft	7
--	---

<i>Lars Jenßen/Simone Dunekacke/Sigrid Blömeke</i> Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis	11
---	----

Berufsbezogene Kompetenzen

<i>Svenja Hammer/Sonja A. Carlson/Timo Ehmke/Barbara Koch-Priewe/ Anne Köker/Udo Ohm/Sonja Rosenbrock/Nina Schulze</i> Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache: Validierung des GSL-Testinstruments	32
--	----

<i>Josef Riese/Christoph Kulgemeyer/Simon Zander/Andreas Borowski/ Hans E. Fischer/Yvonne Gramzow/Peter Reinhold/Horst Schecker/ Elisabeth Tomczyszyn</i> Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik	55
--	----

<i>Simone Dunekacke/Lars Jenßen/Sigrid Blömeke</i> Mathematikdidaktische Kompetenz von Erzieherinnen und Erziehern: Validierung des KomMa-Leistungstests durch die videogestützte Erhebung von Performanz	80
--	----

<i>Franziska Bouley/Stefanie Berger/Sabine Fritsch/Eveline Wuttke/ Jürgen Seifried/Kathleen Schnick-Vollmer/Bernhard Schmitz</i> Der Einfluss von universitären und außeruniversitären Lerngelegenheiten auf das Fachwissen und fachdidaktische Wissen von angehenden Lehrkräften an kaufmännisch-berufsbildenden Schulen	100
--	-----

<i>Olga Zlatkin-Troitschanskaia/Manuel Förster/Susanne Schmidt/ Sebastian Brückner/Klaus Beck</i> Erwerb wirtschaftswissenschaftlicher Fachkompetenz im Studium – Eine mehrbenenanalytische Betrachtung von hochschulischen und individuellen Einflussfaktoren	116
---	-----

Gabriele Kaiser

Erfassung berufsbezogener Kompetenzen von Studierenden.

Ein Kommentar 136

Forschungsbezogene Kompetenzen

Kati Trempler/Andreas Hetmanek mit Christof Wecker/Jan Kiesewetter/

Mia Wermelt/Frank Fischer/Martin Fischer/Cornelia Gräsel

Nutzung von Evidenz im Bildungsbereich – Validierung

eines Instruments zur Erfassung von Kompetenzen

der Informationsauswahl und Bewertung von Studien 144

Sandra Schladitz/Jana Groß Ophoff/Markus Wirtz

Konstruktvalidierung eines Tests zur Messung

bildungswissenschaftlicher Forschungskompetenz 167

Alexandra Winter-Hözl/Kristin Wäschle/Jörg Wittwer/

Rainer Watermann/Matthias Nückles

Entwicklung und Validierung eines Tests zur Erfassung

des Genrewissens Studierender und Promovierender

der Bildungswissenschaften 185

Gabriele Steuer/Tobias Engelschalk/Gregor Jöstl/Anne Roth/

Bastian Wimmer/Bernhard Schmitz/Barbara Schober/Christiane Spiel/

Albert Ziegler/Markus Dresel

Kompetenzen zum selbstregulierten Lernen im Studium:

Ergebnisse der Befragung von Expert(inn)en aus vier Studienbereichen 203

Johannes König

Stand der Forschung zu wissenschaftsbezogenen Kompetenzen

und weiterführende Fragen. Ein Kommentar 226

Sandra Schladitz/Jana Groß Ophoff/Markus Wirtz

Konstruktvalidierung eines Tests zur Messung bildungswissenschaftlicher Forschungskompetenz

Zusammenfassung: Die Befähigung, auf wissenschaftlicher Evidenz basierende Entscheidungen zu treffen, ist zentrales Ziel einer Hochschulausbildung. Das Projekt LeScEd (Learning the Science of Education) integriert Ansätze aus Bereichen wie Informationswissenschaften, Mathematikdidaktik und Psychologie in ein gemeinsames Strukturmodell bildungswissenschaftlicher Forschungskompetenz (BFK). Die aktuelle Studie untersucht Zusammenhänge der BFK zu fluider Intelligenz und selbsteingeschätzter Kompetenz im Sinne diskriminanter bzw. konvergenter Validierung. In Strukturgleichungsmodellen zeigen sich kleine positive Zusammenhänge zur Intelligenz, aber keine zur Selbsteinschätzung. Dies deutet darauf hin, dass BFK mit Intelligenz verwandt, aber von ihr abgrenzbar ist und dass die Selbsteinschätzung keinen geeigneten Indikator für die tatsächliche Kompetenz darstellt.

Schlagworte: bildungswissenschaftliche Forschungskompetenz, Hochschulforschung, Intelligenz, Konstruktvalidität, Strukturgleichungsmodelle

Evidenzbasierte Entscheidungen treffen zu können, ist ein zentrales Qualifizierungsziel eines Hochschulstudiums (KMK, 2005). Voraussetzung hierfür ist unter anderem die Kompetenz, den Forschungsstand des jeweiligen Fachgebietes auf Basis der Literatur strukturieren, verstehen sowie problembezogen evaluieren und reflektieren zu können. Im Bereich der Bildungswissenschaften wird dies als *Educational Research Literacy* (Shank & Brown, 2007; McMillan & Schumacher, 2010) bzw. als bildungswissenschaftliche Forschungskompetenz bezeichnet (BFK; Groß Ophoff, Schladitz, Lohrmann & Wirtz, im Druck). Im Rahmen der evidenzbasierten Beschreibung und Messung von Kompetenzen ist eine differenzierte Konstruktvalidierung insbesondere hinsichtlich konvergenter und diskriminanter Validierung anhand inhaltlich relevanter Referenzkonstrukte erforderlich. Nach Weinert (2001) sollten objektiv erfasste Kompetenzen von u. a. fluider Intelligenz (als domänenübergreifende und nicht veränderbare kognitive Grundfähigkeit) sowie einer subjektiven Kompetenzeinschätzung (als Überzeugung bzgl. der eigenen Fähigkeiten) psychometrisch abgegrenzt werden. Diese Differenzierungen werden im Folgenden auf die BFK übertragen.

1. Theoretischer Hintergrund

Die zunehmende Tendenz zur Bilanzierung von Bildungsprozessen hat sich in den vergangenen Jahren in groß angelegten Schulleistungsstudien wie beispielsweise PISA (Programme for International Student Assessment; Klieme et al., 2010) gezeigt. Die Output-Orientierung gewinnt auch im deutschen Bildungswesen im Hochschulbereich zunehmend an Bedeutung (Zlatkin-Troitschanskaia & Kuhn, 2010). Um Bildungsprozesse evaluieren zu können, ist zunächst ein präzises Verständnis von zugrunde liegenden Kompetenzstrukturen erforderlich (Klieme & Leutner, 2006). Entsprechend wurden in den vergangenen Jahren zahlreiche Projekte ins Leben gerufen, die sich mit der Untersuchung von Kompetenzen im tertiären Bildungsbereich auseinandersetzen (z. B. Teacher Education and Development Study: Mathematics, TEDS-M; Blömeke, Kaiser & Lehmann, 2010). Dabei standen immer häufiger auch Studierende der Bildungswissenschaften, speziell des Lehramts, im Mittelpunkt des Interesses, da diese zukünftigen Generationen kompetent Wissen vermitteln und sie adäquat fördern sollen (Terhart, 2012). AbsolventInnen müssen nicht nur über fachliche und pädagogische Kompetenzen verfügen, sondern auch in der Lage sein, in ihrer pädagogischen Praxis evidenzbasiert Entscheidungen treffen und sich weiterbilden zu können (KMK, 2004). Voraussetzung dafür ist das Grundverständnis, dass zur Beantwortung von Fragestellungen bzw. zur Lösung von Problemen der Bezug auf fundierte wissenschaftliche Erkenntnisse notwendig ist (Brown, Furtak, Timms, Nagashima & Wilson, 2010). Darauf aufbauend ist es wichtig, zielgerichtet Forschungsergebnisse und wissenschaftliche Literatur erschließen, diese kritisch reflektieren sowie daraus handlungsleitende Schlussfolgerungen ableiten zu können (Shank & Brown, 2007).

In der Literatur lassen sich verschiedene empirische Zugänge zur Untersuchung dieser Forschungskompetenz identifizieren, die zwar inhaltliche Bezüge zu den Bildungswissenschaften aufweisen, jedoch vornehmlich durch an diesen Bereich angrenzende Disziplinen geprägt sind (Groß Ophoff et al., im Druck). So beschreiben Brown et al. (2010) für naturwissenschaftliche Kompetenz im schulischen Bereich evidenzbasiertes Schlussfolgern als Prozess, in dem auf eine bestimmte Prämisse bezogene Informationen zunächst analysiert, dann interpretiert und die resultierenden Folgerungen auf eine konkrete Problemstellung angewendet werden. Die am Anfang von Forschungsprozessen stehende Fähigkeit, Informationsbedarf zu erkennen, adäquate Forschungsfragen zu formulieren sowie gezielt und reflektiert zu recherchieren, wird unter dem Schlagwort *Informationskompetenz* untersucht (Catts & Lau, 2008) – beispielsweise indem für ein vorgegebenes Thema angemessene Hypothesen oder semantisch adäquate Schlagwörterkombinationen identifiziert werden müssen. In der Forschung zu *statistischer Kompetenz* wird die Fähigkeit thematisiert, statistische Informationen aus Tabellen, Grafiken oder Ergebnisdarstellungen systematisch analysieren und evaluieren zu können (Ben-Zvi & Garfield, 2004). Die Interpretation solcher Evidenz ermöglicht die Ableitung und Anwendung angemessener Schlussfolgerungen bzw. die kritische Reflexion von Ergebnisinterpretationen, wie sie typischerweise in der Diskussion von Forschungsartikeln anzutreffen sind. Diese Kompetenzfacette wird hauptsächlich als die Fähigkeit zum *kri-*

tischen Denken erforscht (Halpern, 1999). Die drei genannten Kompetenzfacetten Informationskompetenz, statistische Kompetenz und auf Forschung bezogenes kritisches Denken eignen sich nach Groß Ophoff et al. (im Druck) als strukturgebende Dimensionen von BFK.

Um die Qualität eines entsprechenden Testverfahrens beurteilen zu können, ist das Gütekriterium der Validität von entscheidender Bedeutung (Jenßen, Dunekacke & Blömeke, 2015, in diesem Beiheft). So ist über die Analyse der theoretisch postulierten Dimensionalität eines Konstrukts hinaus unter anderem zu prüfen, inwiefern dieses sich von angrenzenden Konstrukten erwartungskonform abgrenzen lässt (Konstruktvalidität; Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Newton & Shaw, 2014). Dazu werden Zusammenhänge mit Merkmalen analysiert, die konzeptuell entweder verwandt (konvergente Validität, z. B. selbsteingeschätzte Kompetenz) oder weiter entfernt sind (diskriminante Validität, z. B. fluide Intelligenz).

Wenngleich die Differenzierung von Kompetenz und Intelligenz nach Weinert (2001) allgemein als erforderlich erachtet wird, wird die konzeptuelle und empirische Trennbarkeit kontrovers diskutiert (Rindermann, 2006). Für die konzeptuelle Unterscheidung sind zwei Aspekte bedeutsam, zum einen die Spezifität und zum anderen die Erlern-/Förderbarkeit. Während sich mit Blick auf die Spezifität Kompetenzen auf klar umrissene Domänen beziehen (z. B. bildungswissenschaftliche Forschung), setzt fluide Intelligenz kein bereichsspezifisches Wissen voraus, sondern beschreibt die generalisierte Fähigkeit, durch Schlussfolgerungen neue kognitive Anforderungen spontan bewältigen zu können (Carpenter, Just & Shell, 1990). Sie wird typischerweise diagnostiziert, indem in figuralen (z. B. CFT 20 R; Weiß, 2006), numerischen und/oder verbalen (z. B. I-S-T 2000 R; Liepmann, Beauducel, Brocke & Amthauer, 2001) Denkaufgaben Regeln erkannt bzw. angewendet werden sollen. Betrachtet man den zweiten Unterscheidungsaspekt, zeigt sich, dass Kompetenz auf bereichsspezifischem Vorwissen basiert, das z. B. im Rahmen von (Hoch-)Schulbildung erworben und gezielt gefördert werden kann. Fluide Intelligenz gilt dagegen als stabiles Persönlichkeitsmerkmal (Neisser et al., 1996), das nicht veränderbar ist. Die Bedeutung der Unterscheidung von BFK und fluider Intelligenz für eine diskriminante Validierung liegt auf der Hand: Nach Brown et al. (2010) kann erst von Evidenz gesprochen werden, wenn Informationen systematisch analysiert und interpretiert werden können, wofür die Fähigkeit schlussfolgernden Denkens grundlegend ist. Zusätzlich sollten aber durch eine universitäre Ausbildung bereichsspezifisch Wissen und Strategien vermittelt werden, welche die Auseinandersetzung mit (bildungswissenschaftlicher) Evidenz unterstützen – dies ist Grundgedanke des sogenannten Value-Added Assessment (Liu, 2011).

Befunde aus Schulleistungsstudien sprechen ebenfalls dafür, Kompetenz und Intelligenz zu unterscheiden. So werden beispielsweise in PISA Kompetenzen als eindeutig abgrenzbar von allgemeinen kognitiven Fähigkeiten konzipiert und entsprechend auch getrennt erhoben (Baumert, Lüdtke, Trautwein & Brunner, 2009; Prenzel, Walter & Frey, 2007). Weiterhin finden sich in der Literatur Hinweise auf mindestens mittlere Zusammenhänge verschiedener Intelligenzfacetten mit Mathematikleistung (Taub, Floyd, Keith & McGrew, 2008) oder Leseleistung (Ramseier & Brühwiler, 2003) – in letzte-

rem Fall selbst unter Berücksichtigung weiterer Variablen wie Geschlecht, Schultyp oder sozial-kultureller Herkunft. Für den Hochschulbereich zeigt sich keine einheitliche Befundlage. So konnte Trapmann (2008) keine Zusammenhänge zwischen Studienerfolg und verbaler bzw. numerischer Verarbeitungskapazität nachweisen, was die Autorin damit erklärt, dass mit fortschreitender Bildung der Zusammenhang zurückzugehen scheint. Dagegen finden sich für figurale Subtests kognitiver Leistungstests kleine bzw. für verbale Subtests mittlere Zusammenhänge mit dem Studienerfolg (Giesen, Gold, Hummer & Jansen, 1986). Substanzielle Zusammenhänge werden dagegen berichtet zwischen fluider Intelligenz und der Fähigkeit zum komplexen Problemlösen, bei dem angesichts einer Vielzahl interagierender Informationen Komplexität reduziert werden muss (Süß, 2003).

Während die Analyse des Zusammenhangs zwischen Kompetenz und Intelligenz eine Form diskriminanter Validierung darstellt, ermöglicht die Analyse des Zusammenhangs zwischen unterschiedlichen Erhebungsmethoden des gleichen Konstrukts, also z.B. objektiver Testleistung und Selbsteinschätzung, die Prüfung konvergenter Validität. Die Selbsteinschätzung eigener Fähigkeiten ist definiert als die Wahrnehmung der Person ihrer aktuellen Kompetenz bezüglich einer bestimmten Tätigkeit (Wigfield & Eccles, 2000). Im Sinne des Multitrait-Multimethod-Ansatzes (Campbell & Fiske, 1959) wären – im Gegensatz zur Abgrenzung von fluider Intelligenz – hohe Zusammenhänge zu objektiv gemessener Kompetenz zu erwarten. Allerdings ist speziell die Befundlage in diesem Bereich kontrovers: So wurden in schulischen Stichproben allenfalls kleine bis mittlere Zusammenhänge zwischen den Erhebungsmethoden gefunden (Freiberger, Steinmayr & Spinath, 2012). Auch Studierende scheinen sich bzgl. ihrer Kompetenz zu verschätzen (Chevalier, Gibbons, Thorpe, Snell & Hoskins, 2009).

Basierend auf den dargestellten Befunden stellt dieser Beitrag Analysen zur Konstruktvalidität eines Testinstruments zur Erfassung von BFK vor, das im Verbundprojekt Learning the Science of Education (Schladitz et al., 2013) entwickelt wurde. Es wird erwartet, dass die BFK

- a) im Sinne diskriminanter Validierung kleine bis mittlere Zusammenhänge zur fluiden Intelligenz sowie
- b) im Sinne konvergenter Validierung mittlere bis hohe Zusammenhänge zur selbsteingeschätzten Kompetenz aufweist.

Entsprechend der inhaltlichen Ausrichtung der Kompetenzfacetten sowie der Konzeption der Intelligenzmaße und der Skala zur Selbsteinschätzung werden unterschiedlich starke Effekte erwartet. Die Zusammenhänge zwischen der sprachfreien Messung der Intelligenz und den Kompetenzfacetten sollten in allen Fällen im mittleren Bereich liegen. Für die verbale Messung der Intelligenz wird erwartet, dass mittlere Zusammenhänge zu denjenigen Kompetenzfacetten bestehen, die einen hohen Anteil sprachlicher Informationen beinhalten (Informationskompetenz, forschungsbezogenes kritisches Denken), und kleine Zusammenhänge zu statistischer Kompetenz.

Die neu entwickelte Skala zur selbsteingeschätzten BFK sollte dieselben Facetten abbilden wie der Kompetenztest. Entsprechend wird erwartet, dass zwischen den jeweils gleichen Facetten höhere Zusammenhänge bestehen als zwischen den nicht zusammengehörigen Facetten. Im Sinne diskriminanter Validierung werden keine Zusammenhänge zwischen selbsteingeschätzter BFK und fluider Intelligenz erwartet.

2. Methoden

2.1 Instrumente

Der folgende Abschnitt stellt die für diese Studie relevanten Instrumente vor. Zusätzlich wurden noch weitere Skalen eingesetzt (z. B. Motivationsskalen), die für den vorliegenden Beitrag nicht relevant sind.

Kompetenztest

Das Instrument zur Erfassung der BFK wurde im Rahmen des Projekts neu entwickelt. Anhand bestehender Literatur im Bereich Forschungskompetenz wurden die Inhaltsbereiche Informationskompetenz, statistische Kompetenz sowie forschungsbezogenes kritisches Denken als relevante Facetten identifiziert und als Konstruktionsheuristik für die Itementwicklung genutzt (Groß Ophoff et al., im Druck). Zur Gewährleistung der Inhaltsvalidität wurden die ersten Aufgabenentwürfe mit ExpertInnen ($N = 5$) auf dem Gebiet der Kompetenzmessung bzw. Pädagogischen Psychologie diskutiert. Nach einer Voruntersuchung ($N = 6$ Studierende) bezüglich der Verständlichkeit der Items bestand der Itempool aus 226 Einzelitems (Beispiele s. Abb. 1).

Die postulierte Multidimensionalität und die angestrebten Skalenreliabilitäten bedingten eine sehr große Itemanzahl, weshalb für die Datenerhebung ein unvollständiges Blockdesign (Frey, Hartig & Rupp, 2009) verwendet wurde. Die Items wurden auf Blöcke zu je acht Aufgaben aufgeteilt. Diese wurden so zusammengestellt, dass die Bearbeitungszeit etwa gleich war, alle Facetten auf Aufgabenebene gleichmäßig vertreten und keine Aufgaben mit dem gleichen Itemstamm in einem Block enthalten waren, um Redundanzen und lokale Abhängigkeiten zu vermeiden. Jeweils vier Aufgabenblöcke wurden zu einem Testheft zusammengefasst (Abb. 2), sodass jeder Person 32 Testaufgaben vorlagen. Um Reihenfolgeeffekte einzuschränken, wurde jedes Testheft zur Hälfte in blockinverser Reihenfolge dargeboten.

Es wurden 23 Items zu Informationskompetenz, 68 zu statistischer Kompetenz und 91 zu forschungsbezogenem kritischem Denken in die Analysen einbezogen. Die Diskrepanz zur ursprünglichen Itemanzahl liegt darin begründet, dass nur Items im Multiple-Choice- und Fill-in-Format in die aktuelle Analyse aufgenommen wurden sowie Items mit nicht adäquater Modellpassung im Vorfeld ausgeschlossen wurden (Groß Ophoff et al., im Druck).

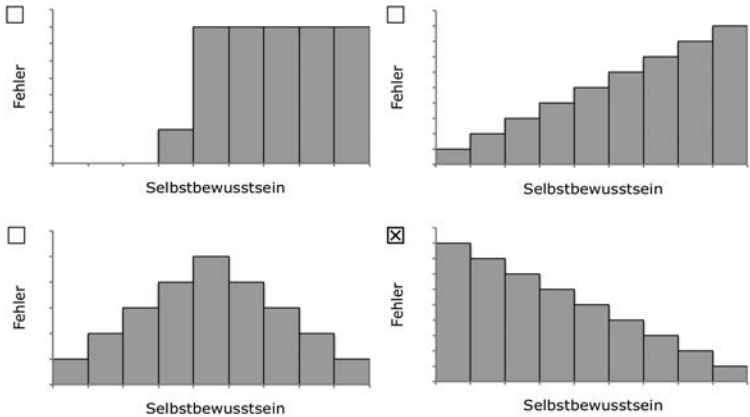
Um bei einer Datenbankrecherche Suchwörter miteinander zu verbinden, können die logischen Operatoren AND, OR bzw. NOT verwendet werden, die zu unterschiedlichen Suchergebnissen führen.

Bringen Sie die logischen Kombinationen mit den Zahlen 1 (= meiste Suchergebnisse) bis 3 (wenigste Suchergebnisse) in die richtige Reihenfolge.

- a) ___ Heterogenität
- b) ___ Heterogenität OR Grundschule
- c) ___ Heterogenität NOT Grundschule

In Experimenten hat sich gezeigt, dass ein höheres Selbstbewusstsein die Fehleranzahl in Diktaten verringert.

Welche der folgenden Verteilungsformen bildet dieses Ergebnis ab?



Sie lesen die folgenden beiden Kurzdarstellungen von Forschungsbefunden:

Befund A: Die Notwendigkeit einer grundlegenden Reform des Bildungssystems wird durch die nachgewiesene hohe Unzufriedenheit der untersuchten Berufsgruppen (Lehrer/innen, Pädagog/inn/en, Erzieher/innen) belegt. Zukünftig müssen die Struktur und die Prozesse in Bildungseinrichtungen stärker auf die Bedürfnisse der beteiligten Berufsgruppen abgestimmt werden, da eine Verbesserung der Ausbildungsqualität unbedingt erforderlich ist.

Befund B: Der im Gruppenvergleich nachgewiesene positive Effekt der Bildung von Lehrerteams auf die Unterrichtsqualität unterstützt die Annahme, dass durch Ausbildungsteams positive Effekte erzielt werden können. Zukünftig sollte die Entwicklung und Prüfung der Effekte konkreter Teambildungsmaßnahmen in Schulen weiter untersucht werden.

Geben Sie an, welche Merkmale eher für Befund A bzw. Befund B zutreffen:

	Eher für A zutreffend	Für A und B in ähnlichem Maß zutreffend	Eher für B zutreffend
Die Ableitungen der Studie sind direkt auf die Untersuchungsergebnisse bezogen.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Die untersuchten Merkmale werden konkret benannt.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Die verwendete Untersuchungsanlage wird deutlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Abb. 1: Beispielitems zu den Kompetenzfacetten Informationskompetenz (oben), statistische Kompetenz (Mitte) und forschungsbezogenes kritisches Denken (unten)

Testheft	1		2		3		...	18		19		20	
	V	R	V	R	V	R		V	R	V	R	V	R
	1	4	2	5	3	6		18	1	19	2	20	3
	2	3	3	4	4	5		19	20	20	1	1	2
	3	2	4	3	5	4		20	19	1	20	2	1
	4	1	5	2	6	3		1	18	2	19	3	20

Abb. 2: Verteilung der Aufgabenblöcke 1–20 á 8 Aufgaben im Testheftdesign mit 20 Testheften, 4 Blockpositionen pro Testheft; V = Vorwärtsreihenfolge der Blöcke, R = Rückwärtsreihenfolge

Intelligenztests

Zur Erfassung der fluiden Intelligenz wurden Subskalen aus zwei verschiedenen Tests eingesetzt. Die Subskala „Gemeinsamkeiten finden“ aus dem Intelligenz-Struktur-Test (I-S-T 2000 R; Liepmann et al., 2001) erfasst verbale Intelligenz. In den für die Studie relevanten Alterskohorten werden im Skalenhandbuch folgende Mittelwerte und Standardabweichungen für die aus 20 Items bestehende Subskala angegeben: 15–20 Jahre: 10.72 ($SD = 3.49$); 21–25 Jahre: 11.12 ($SD = 3.67$); 26–30 Jahre: 10.39 ($SD = 3.92$); 31–40 Jahre: 9.83 ($SD = 3.98$). Zusätzlich wurde die Subskala „Reihenfortsetzen“ aus dem Culture Fair Intelligence Test (CFT 20 R; Weiß, 2006) eingesetzt, um neben der sprachabhängigen (= verbalen) auch eine sprachfreie Messung fluiden Intelligenz zu erhalten. Je einem Drittel der Personen wurde der CFT 20 R oder der I-S-T 2000 R zur Bearbeitung vorgelegt, sodass jede/r TeilnehmerIn nur einen der Tests bearbeitet hat. Der CFT 20 R bietet neben den Original-Subskalen jeweils auch eine Parallelversion (Testform A und B). Um die Testzeit für alle Personen gleichzuhalten, wurden hier beide Parallelversionen der Subskala verwendet.

Skala zur selbsteingeschätzten BFK

Es wurde eine Skala mit acht Items neu entwickelt, die sich auf Aspekte bezieht, die auch im Kompetenztest gemessen werden: Literaturrecherche/Fragenstellen (vgl. Informationskompetenz; Items s. Tab. 3), Umgang mit Methoden-/Ergebnisdarstellungen (vgl. statistische Kompetenz) sowie kritische Reflexion von Befunden (vgl. forschungsbezogenes kritisches Denken). Zusätzlich wurde ein Globalitem zum Verständnis bildungswissenschaftlicher Studien gestellt. Die Einschätzung erfolgte auf einer fünfstufigen Likert-Skala.

2.2 Stichprobe

Es konnten 1360 Personen an sechs deutschen Hochschulen für die Teilnahme gewonnen werden. Die Merkmale der (Teil-)Stichproben sind in Tabelle 1 aufgeführt.

	Gesamtstichprobe	Teilstichprobe „Gemeinsamkeiten finden“ (verbale Intelligenz)	Teilstichprobe „Reihenfortsetzen“ (sprachfreie Intelligenz)
<i>N</i>	1360	461	434
Alter (<i>M</i> , <i>SD</i>)	22.93 (3.95)	22.74 (3.70)	22.88 (3.30)
Semester (Median)	3	3	3
weiblich (%)	75.4	75.6	75.5
Studiengänge (%)			
Lehramt	61.8	63.7	61.8
Erziehungswissenschaft	23.1	22.7	22.7
Gesundheitspädagogik	5.1	4.5	5.5
andere	8.9	7.7	9.0

Anmerkungen. *N* = Stichprobengröße; *M* = Mittelwert; *SD* = Standardabweichung.

Tab. 1: Deskriptive Statistiken der Gesamt- und Teilstichproben

2.3 Analyseverfahren

Basis für die Zusammenhangsanalysen sind Skalen, die den gängigen psychometrischen Gütekriterien genügen (Adams, 2002; Schermelleh-Engel, Moosbrugger & Müller, 2003). In diesem Sinne wurden zunächst die BFK (Groß Ophoff, Schladitz & Wirtz, 2014) sowie die Intelligenzmaße und die selbsteingeschätzte Kompetenz skalenanalytisch evaluiert. Für jede der Kompetenzfacetten (Informationskompetenz, statistische Kompetenz, forschungsbezogenes kritisches Denken) wurden fünf Personenschätzer (Plausible Values; von Davier, Gonzalez & Mislevy, 2009) ermittelt, die als Kriterium für die vorgestellten strukturanalytischen Analysen dienten. Im Rahmen der vorliegenden Analyse wurde für die dichotomen Daten aus den Intelligenzskalen auf Messmodellebene die Passung zu den Annahmen des Raschmodells (1PL-Modell) und des Birnbaum-Modells (2PL-Modell) auf Basis der Informationskriterien Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) sowie Consistent Akaike Information Criterion (CAIC) kontrastierend geprüft. Niedrigere Werte zeigen hierbei eine bessere Datenkompatibilität unter Berücksichtigung der Modellkomplexität an (Schermelleh-Engel et al., 2003). Bei der anschließenden Überprüfung der Itemeigenschaften wurden Items beibehalten, die anhand des Weighted Mean Square (WMNSQ) und des dazugehörigen t-Werts gut zur Gesamtskala passten. Die Werte sollten im Bereich $0.80 \leq \text{WMNSQ} \leq 1.20$ liegen und keine Signifikanz aufweisen. Diese Werte wurden von Linacre (1994) als stichprobengrößenfaire Grenzwerte vorgeschlagen, um größere Stichprobenumfänge nicht zu bestrafen. Die EAP/PV (expected a posteriori/plausible value)-Reliabilität, die mit Cronbach's Alpha vergleichbar ist, gibt die Messgenauigkeit der Skala an.

Zur Analyse der kontinuierlichen Daten aus der Skala zur selbsteingeschätzten Kompetenz wurde zunächst mit einem explorativen Strukturgleichungsmodell (Asparouhov & Muthén, 2009) die Faktorenstruktur analysiert und anschließend die am besten zu den Daten passende Struktur konfirmatorisch modelliert. Die Modellgüte wurde anhand der gängigen Maße überprüft, wonach der Root Mean Square Error of Approximation (RMSEA) höchstens .08 und der Comparative-Fit-Index (CFI) mindestens .95 betragen sollte (Schermelleh-Engel et al., 2003).

Nach Sicherstellung aller Skalengütekriterien wurden die Zusammenhänge zwischen Kompetenzfacetten, Intelligenzkonstrukten sowie selbsteingeschätzter Kompetenz analysiert und ebenfalls anhand der o. g. Gütekriterien bewertet.

3. Ergebnisse

3.1 Psychometrische Prüfung der Intelligenzskalen

Der Modellvergleich für die Intelligenzskalen ergab für beide Konstrukte eine bessere Passung des 2PL-Modells gegenüber dem Rasch-Modell (Tab. 2). Aus diesem Grund wurde für alle weiteren Analysen das 2PL-Modell verwendet.

Betrachtet man die psychometrischen Eigenschaften der einzelnen Items, zeigen sich für die Skala „Gemeinsamkeiten finden“ aus dem Test zur verbalen Intelligenz Lösungshäufigkeiten zwischen 4.99 % und 93.26 % mit einer durchschnittlichen Lösungshäufigkeit von 57.60 %. Die Skala ist so konzipiert, dass die Aufgaben im Verlauf der Skala schwieriger werden. Dies bildet sich auch im vorliegenden Fall ab: In der ersten Hälfte der Skala liegt die durchschnittliche Lösungshäufigkeit bei 78.64 %, in der zweiten Hälfte bei 36.55 %. Die Werte des WMNSQ liegen mit [0.99; 1.05] durchgehend im guten Bereich und weisen keine Signifikanz auf. Die aus dem 2PL-Modell ermittelten Schwierigkeiten der Items liegen im Bereich [−7.12; 23.85] mit Diskriminationsparametern von [0.12; 3.25]. Die EAP/PV-Reliabilität für diese Skala kann mit .73 als zufriedenstellend angesehen werden.

	„Gemeinsamkeiten finden“ (verbale Intelligenz)		„Reihenfortsetzen“ (sprachfreie Intelligenz)	
	1 PL	2 PL	1 PL	2 PL
AIC	9058.651	8887.220	9161.037	9046.856
BIC	9145.452	8056.690	9275.082	9270.873
CAIC	9166.452	9097.690	9303.082	9325.873

Anmerkungen. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; CAIC = Consistent Akaike Information Criterion. Das jeweils am besten passende Modell wurde hervorgehoben.

Tab. 2: Modellvergleiche für die Intelligenzskalen

Der Mittelwert korrekt gelöster Items unterschied sich mit 11.52 ($SD = 2.97$) von den Normwerten der Skala „Gemeinsamkeiten finden“ aus dem I-S-T 2000 R. So lagen Studierende in der Altersgruppe von 15 bis 20 Jahren ($n = 104$) mit einem Mittelwert von 11.93 (95 %-Konfidenzintervall: 11.44; 12.42, vgl. 3.1) signifikant über dem der Referenzgruppe und streuten entsprechend der Standardabweichung von 2.56 (95 %-Konfidenzintervall: 2.21; 2.91) signifikant geringer. Die größte Altersgruppe der 21- bis 25-Jährigen ($n = 301$) unterschied sich mit einem Mittelwert von 11.51 nicht von der Referenzgruppe, erwies sich aber mit einer Standardabweichung von 2.97 (95 %-Konfidenzintervall: 2.74; 3.21) ebenfalls als signifikant homogener.

Aus der Skala „Reihenfortsetzen“ sind die WMNSQ-Werte mit [0.89; 1.09] allesamt zufriedenstellend und nicht signifikant. Die durchschnittliche Lösungshäufigkeit liegt bei 78.24 % mit Werten im Bereich [32.03 %; 97.23 %]. Auch in diesem Fall zeigt sich die in der Testentwicklung intendierte Steigerung der Aufgabenschwierigkeit über die Skala hinweg. Halbiert man die beiden Testformen A und B, ergeben sich für die ersten Hälften durchschnittliche Lösungshäufigkeiten von 89.95 % bzw. 91.67 % und für die zweiten Hälften von 76.27 % bzw. 51.50 %. Die ermittelten Schwierigkeiten liegen damit korrespondierend im Bereich $[-2.60; 0.83]$, die Diskriminationsparameter zwischen [0.56; 5.35]. Insgesamt weist die Skala mit $EAP/PV = .71$ eine zufriedenstellende Reliabilität auf.

3.2 Struktur der Itemgruppen zur „selbsteingeschätzten Kompetenz“

Die Analyse des explorativen Strukturgleichungsmodells lieferte Hinweise auf eine zweidimensionale Struktur (Tab. 3). Den ersten Faktor bilden die vier Items aus den Bereichen Recherche/Fragenstellen und kritische Reflexion, die drei Items des zweiten Faktors beziehen sich auf das Lesen von Methoden-/Ergebnisdarstellungen. Das Item zum globalen Verständnis wurde ausgeschlossen. Die beiden Dimensionen weisen jeweils zufriedenstellende Reliabilität auf ($\alpha = .76$ bzw. $\alpha = .70$) und korrelieren latent mit .70 miteinander. Die Gütekriterien für das konfirmatorische Strukturmodell liegen ebenfalls im akzeptablen Bereich ($RMSEA = .066$; $CFI = .970$).

3.3 Zusammenhang zwischen Kompetenzfacetten, Intelligenzkonstrukten und selbsteingeschätzter Kompetenz

In Strukturgleichungsmodellen wurden anschließend die Zusammenhänge zwischen Kompetenzfacetten, Intelligenzkonstrukten und der selbsteingeschätzten Kompetenz analysiert (s. Abb. 3, aus Gründen der Übersichtlichkeit sind die Messmodelle nicht abgebildet). Um die lokale Abhängigkeit der Schätzungen aus denselben Simulationsläufen zu berücksichtigen, wurden simulationsspezifische Residualkorrelationen der manifesten Plausible Values definiert. Als Schätzalgorithmus wurde Mean and Variance Adjusted Weighted Least Squares (WLSMV; z.B. Beauducel & Herzberg, 2006) ver-

Item	λ	SE	R^2
<i>Dimension 1 – Literaturrecherche/Fragenstellen & kritische Reflexion</i>			
Ich fühle mich sicher bei der Literaturrecherche für wissenschaftliche Arbeiten.	.47	.05	.24
Ich fühle mich sicher in der Formulierung wissenschaftlicher Fragestellungen und Hypothesen.	.74	.05	.52
Ich kann Forschungsergebnisse kritisch reflektieren.	.66	.05	.47
Ich kann die Qualität einer bildungswissenschaftlichen Studie sicher beurteilen.	.59	.03	.34
<i>Dimension 2 – Lesen von Methoden-/Ergebnisdarstellungen</i>			
Ich kann Daten aus Diagrammen, Tabellen und Texten gut erfassen und sicher interpretieren.	.70	.05	.46
Ich kann methodische und statistische Aspekte aus bildungswissenschaftlichen Studien gut verstehen.	.80	.03	.64
Ich kann Ergebnisse einer bildungswissenschaftlichen Studie angemessen interpretieren.	.55	.04	.50

Anmerkungen. λ = standardisierte Faktorladung; SE = Standardfehler; R^2 = Varianzaufklärung. Für alle Ladungen gilt $p < .001$.

Tab. 3: Faktorladungen, Fehlerterme und Varianzaufklärung der Items zur selbsteingeschätzten BFK

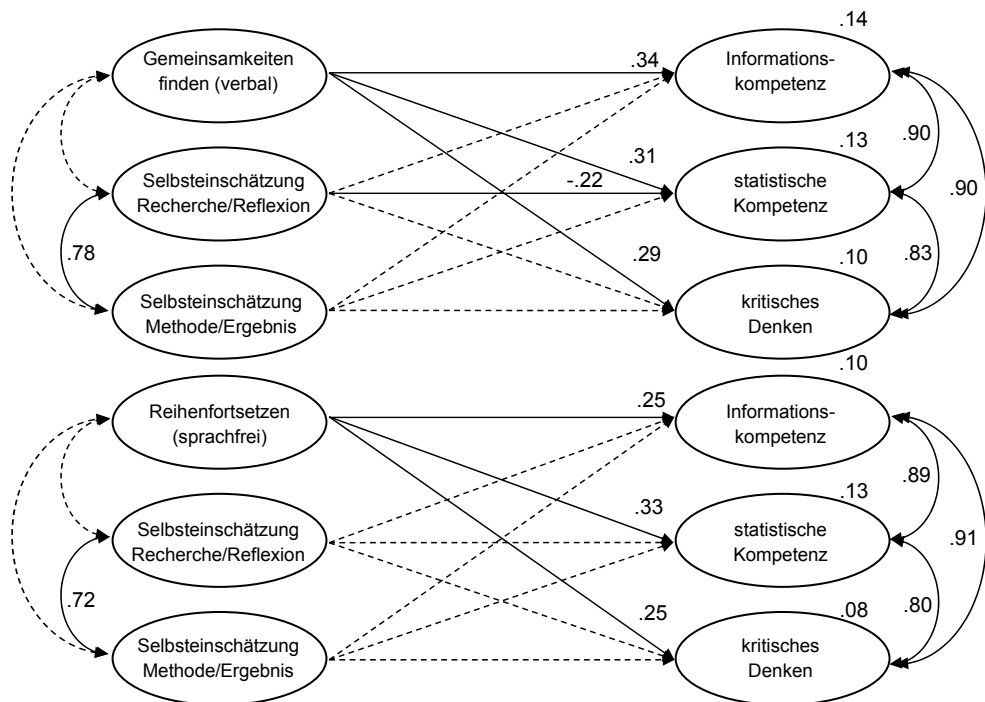


Abb. 3: Strukturmodelle zum Zusammenhang zwischen den Intelligenzskalen, der selbsteingeschätzten Kompetenz und den Kompetenzfacetten. Dargestellt sind signifikante standardisierte Regressionsgewichte, latente Korrelationen sowie die Varianzaufklärung. Nicht-signifikante Pfade sind gestrichelt dargestellt.

wendet, wobei fehlende Werte listenweise ausgeschlossen wurden. Die Maße der Modellgüte waren in beiden Fällen gut bis zufriedenstellend (Skala „Gemeinsamkeiten finden“: RMSEA = .020; CFI = .966; Skala „Reihenfortsetzen“: RMSEA = .019; CFI = .962). Die dargestellten vollständig standardisierten Regressionsgewichte lagen für die Intelligenzkonstrukte im schwachen bis mittleren Bereich (Cohen, 1988). Die Effekte der selbsteingeschätzten Kompetenz erwiesen sich hingegen bis auf einen Fall als nicht bedeutsam ($|\beta| \leq .18$).

4. Diskussion

Da dieser Beitrag mit dem Ziel konvergenter und diskriminanter Validierung der BFK verfasst wurde, lag der Fokus auf den Zusammenhängen mit fluider Intelligenz sowie der Selbsteinschätzung der Forschungskompetenz. Die Darstellung der detaillierten Skalenanalyse und Struktur der BFK ist für einen weiteren Beitrag vorgesehen (erste Ergebnisse bei Groß Ophoff et al., im Druck), ebenso wie Befunde zu Korrelationen mit weiteren, im Rahmen des Forschungsprojekts LeScEd erhobenen Konstrukten.

Die vorangehende Überprüfung der Struktur der Intelligenzskalen zeigt, dass die Analyse mittels IRT-Methoden eine wertvolle Ergänzung zur ursprünglichen Konstruktion anhand klassischer Testtheorie darstellt. So konnte an vergleichsweise großen Stichproben verdeutlicht werden, dass die ermittelten Reliabilitäten grenzwertig zufriedenstellend sind und durch weitere Itemanalysen ggf. noch verbessert werden können. Das besser zu den Daten passende 2PL-Modell hat gegenüber dem Rasch-Modell den Nachteil, dass die identifizierten Dimensionen schwieriger zu interpretieren sind (Rauch & Hartig, 2012). Einschränkend ist außerdem zu berücksichtigen, dass die hier untersuchte studentische Stichprobe z. T. signifikant höhere Mittelwerte und eine geringere Varianz in der Skala „Gemeinsamkeiten finden“ der verbalen Intelligenz aufweist als die im Manual wiedergegebenen Normierungsstichproben (vgl. 3.1; Liepmann et al., 2001), was zu einer Verzerrung der Effektschätzungen führen kann (Urban & Mayerl, 2008). Eine erneute Überprüfung der Intelligenzskalen anhand probabilistischer Testmodelle und der Passung des Rasch-Modells in weniger spezifischen Stichproben sollte daher in Erwägung gezogen werden.

Die entwickelte Skala zur selbsteingeschätzten BFK zeigt vielversprechende Gütekriterien. Man muss allerdings berücksichtigen, dass die Struktur nicht deckungsgleich mit den Facetten des Kompetenztests ist. In der Analyse der Skalenstruktur bildeten sich zwei statt drei Inhaltsbereiche ab. Es scheint, als gäbe es Unterschiede hinsichtlich der Verarbeitung hauptsächlich verbaler (Informationskompetenz, forschungsbezogenes kritisches Denken) und hauptsächlich statistischer Informationen (statistische Kompetenz). Die Nähe zu Kompetenzkonstrukten aus der Schulleistungsforschung (Lesekompetenz, mathematische Kompetenz) liegt auf der Hand und sollte in künftigen Studien explizit berücksichtigt werden.

4.1 Diskriminante und konvergente Validierung

Die Zusammenhänge zwischen der BFK und der Intelligenz (diskriminante Validierung) liegen wie erwartet im mittleren Bereich. Die somit schwächeren Zusammenhänge, als sie z. B. in PISA ermittelt wurden (Baumert et al., 2009; Prenzel et al., 2007), gehen einher mit Trapmanns (2008) Standpunkt, dass der Zusammenhang zwischen Intelligenz und akademischer Leistung im Verlauf des Bildungsweges abnimmt. Es zeigt sich also, dass BFK und fluide Intelligenz (mit den Subskalen „Gemeinsamkeiten finden“ und „Reihenfortsetzen“) zwar verwandte, aber empirisch gut voneinander abgrenzbare Konstrukte darstellen. Die auf Vorwissen und Lernerfahrungen basierende Forschungskompetenz stellt offensichtlich ein über die fluide Intelligenz hinausgehendes Merkmal dar, was für eine getrennte Erfassung beider Bereiche spricht.

Aufseiten der selbsteingeschätzten BFK (konvergente Validierung) zeigen sich dagegen unerwartet fast keine Zusammenhänge zur objektiv gemessenen BFK. Die einzige signifikante Korrelation erweist sich sogar als negativ, wonach eine selbsteingeschätzte höhere Kompetenz in Recherche und Reflexion von Evidenz mit einer geringeren statistischen Kompetenz einhergeht. Dies kann ein Hinweis darauf sein, dass mit einer allgemeinen, also nicht auf ein spezifisches Lehrangebot bezogenen Selbsteinschätzung von Kompetenz eher Präferenzen für bestimmte Aspekte der Beschäftigung mit bildungswissenschaftlicher Forschung gemessen werden – wobei der Effekt eher klein ausfällt und nicht überinterpretiert werden sollte. Die nicht aufzeigbaren Zusammenhänge zwischen objektiver und subjektiver BFK können darin begründet sein, dass die selbsteingeschätzte Kompetenz tatsächlich keinen geeigneten Prädiktor für die objektive Kompetenzausprägung darstellt. Damit bestätigt sich die bereits angedeutete Diskrepanz zwischen Validierungstheorie (Campbell & Fiske, 1959) und empirischer Befundlage (z. B. Chevalier et al., 2009): Trotz Messung desselben Konstrukts – was für einen hohen Zusammenhang spräche – scheinen die beiden Operationalisierungen voneinander unabhängig. Eine generelle Schwierigkeit beim Einsatz von Selbsteinschätzungsskalen ist deren Losgelöstheit von tatsächlichen Kompetenzaufgaben. So ist denkbar, dass sich bei Vorgabe der spezifischen Kompetenzitems mit der Frage nach der selbst eingeschätzten Lösungswahrscheinlichkeit andere Ergebnisse zeigen würden. Die abstrakte Bewertung recht weit formulierter Kompetenzen bietet evtl. ein hohes Potential für eine falsche Einschätzung des eigenen Kompetenzniveaus. Es empfiehlt sich daher, für eine zuverlässige Kompetenzerfassung nicht ausschließlich Instrumente zur Selbsteinschätzung (z. B. Braun, Gusy, Leidner & Hannover, 2008) einzusetzen, sondern, wenn möglich, eine objektive Kompetenzmessung zu ermöglichen.

Trotz der nachweislich besser passenden dreidimensionalen Kompetenzstruktur (Groß Ophoff et al., 2014) zeigen sich kaum differenzierte Zusammenhänge zwischen Intelligenz und Selbsteinschätzung auf der einen sowie den drei Kompetenzfacetten auf der anderen Seite. In Verbindung mit den sehr hohen Interkorrelationen der Facetten wirft dieser Umstand die Frage auf, inwiefern es praktisch sinnvoll (z. B. für die Rückmeldung an Studierende im Rahmen von Lehrveranstaltungen) ist, BFK so aus-

fürlich zu differenzieren. Auch wenn z. T. in Large-Scale-Studien trotz vergleichbarer oder sogar höherer Zusammenhänge eine Trennung von Teilkompetenzen als Beitrag zur grundlegenden Erforschung des Konstrukts beibehalten wird (Artelt & Schlagmüller, 2004), sollten vertiefende Analysen zur weiteren Klärung beitragen. Letztlich bleibt es vermutlich eine inhaltliche, mit Blick auf den Verwendungszusammenhang (i. S. von Consequential Validity; vgl. Newton & Shaw, 2014) zu treffende Entscheidung für oder gegen die Unterscheidung der Kompetenzfacetten.

Abschließend lässt sich festhalten, dass die diskriminante Validierung der BFK anhand der fluiden Intelligenz vorgenommen werden konnte. Mit der selbsteingeschätzten Kompetenz im Sinne konvergenter Validierung dagegen wird offensichtlich nicht das gleiche Konstrukt, sondern eventuell individuelle Präferenzen bei der Auseinandersetzung mit bildungswissenschaftlicher Forschung gemessen.

4.2 Ausblick auf zukünftige Forschung

Der Einbezug zweier Skalen aus verschiedenen Intelligenzbereichen (verbal und sprachfrei) trägt nicht zur weiteren Validierung bzw. Aufklärung des Konstrukts BFK bei. Jedoch lässt die aus testökonomischen Gründen vorgenommene Aufteilung der Intelligenzskalen auf zwei nichtüberlappende Teilstichproben keine Analyse inkrementeller Effekte der spezifischen Intelligenzarten zu. Eine gemeinsame Erfassung in einer Stichprobe sowie der Einbezug anderer Intelligenzkomponenten (z. B. numerische Intelligenz) können weitere Hinweise auf differenzierte Zusammenhangsstrukturen zur BFK liefern. Weiterhin werden differenzielle Itemeigenschaften (Zumbo, 2007) in Abhängigkeit von der Intelligenz untersucht, um genauere Informationen darüber zu erhalten, welche Items bzw. Itemgruppen besonders stark von der Intelligenzausprägung beeinflusst werden.

Neben der Intelligenz und der Selbsteinschätzung können sich weitere Personenmerkmale wie z. B. Motivation (z. B. Spinath & Steinmayr, 2012) als relevant für die Ausbildung der BFK erweisen. Weinert (2001) empfiehlt auch für solche nicht-kognitiven Merkmale eine getrennte Erfassung, um Zusammenhänge zur jeweiligen Kompetenz analysieren zu können. Darüber hinaus spielen vor allem in einem – im Gegensatz zur schulischen Ausbildung – relativ frei gestaltbaren Hochschulstudium Lerngelegenheiten (Blömeke, Suhl, Kaiser, Felbrich & Schmotz, 2010) eine Rolle. Diese Aspekte werden derzeit im Rahmen einer längsschnittlichen Untersuchung der BFK analysiert.

Die Befunde zur Abgrenzbarkeit der BFK von fluider Intelligenz bieten eine Grundlage für die Entwicklung gezielter Lehr- bzw. Weiterbildungsangebote. So können je nach Zielsetzung der Lehrveranstaltung ausgewählte Teilkompetenzen gefördert werden, die zwar mit allgemeiner Intelligenz zusammenhängen, aber auch darüber hinausgehend zu einer verbesserten Leistung im Umgang mit bildungswissenschaftlicher Forschung führen. Jedoch bleibt das Desiderat bestehen, inwiefern die Vermittlung entsprechender Kompetenzfacetten tatsächlich evidenzbasiertes Arbeiten und kontinuierliche Professionalisierung in der späteren beruflichen Praxis unterstützen kann. Hier

bieten sich Studien an, die den Übergang in den Beruf verfolgen (z. B. COACTIV-R; Kunter et al., 2011), die Kompetenzen Berufstätiger systematisch erfassen und idealerweise auch langfristig begleiten.

Literatur

- Adams, R. J. (2002). Scaling PISA cognitive data. In R. J. Adams & M. L. Wu (Hrsg.), *PISA 2000 technical report* (S. 99–108). Paris: OECD.
- Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen? Dimensionsanalysen und Ländervergleiche. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 169–196). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397–438.
- Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review*, 4(3), 165–176 [DOI: 10.1016/j.edurev.2009.04.002].
- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of Maximum Likelihood versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(2), 186–203 [DOI: 10.1207/s15328007sem1302_2].
- Ben-Zvi, D., & Garfield, B. (2004). *The challenge of developing statistical literacy, reasoning and thinking*. New York: Kluwer Academic Publishers.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- Blömeke, S., Suhl, U., Kaiser, G., Felbrich, A., & Schmotz, C. (2010). Lerngelegenheiten und Kompetenzerwerb angehender Mathematiklehrkräfte im internationalen Vergleich. *Unterrichtswissenschaft*, 38(1), 29–50.
- Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Kompetenzorientierte Lehrevaluation – Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, 54(1), 30–42.
- Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The Evidence-Based Reasoning Framework: Assessing scientific reasoning. *Educational Assessment*, 15(3/4), 123–141.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the Multi-trait-Multimethod Matrix. *Psychological Bulletin*, 56(2), 81–105.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404–431.
- Catts, R., & Lau, J. (2008). *Towards information literacy indicators: Conceptual framework paper*. Paris.
- Chevalier, A., Gibbons, S., Thorpe, A., Snell, M., & Hoskins, S. (2009). Students' academic self-perception. *Economics of Education Review*, 28(6), 716–727.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale: Lawrence Erlbaum Associates.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), S. 281–302.

- Freiberger, V., Steinmayr, R., & Spinath, B. (2012). Competence beliefs and perceived ability evaluations: How do they contribute to intrinsic motivation and achievement? *Learning and Individual Differences*, 22, 518–522 [DOI: 10.1016/j.lindif.2012.02.004].
- Frey, A., Hartig, J., & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice*, 28(3), 39–53.
- Giesen, H., Gold, A., Hummer, A., & Jansen, R. (1986). *Prognose des Studienerfolgs*. Frankfurt a.M.: Institut für Pädagogische Psychologie.
- Groß Ophoff, J., Schladitz, S., Lohrmann, K., & Wirtz, M. (im Druck). Evidenzorientierung in bildungswissenschaftlichen Studiengängen: Entwicklung eines Strukturmodells zur Forschungskompetenz. In W. Bos, K. Drossel & R. Strietholt (Hrsg.), *Empirische Bildungsforschung und evidenzbasierte Reformen im Bildungswesen*. Münster: Waxmann.
- Groß Ophoff, J., Schladitz, S., & Wirtz, M. (2014, März). *Struktur von Research Literacy in den Bildungswissenschaften*. Vortrag auf der Tagung der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE). Berlin.
- Halpern, D. F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning*, 1999, 69–74 [DOI: 10.1002/tl.8005].
- Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. *Zeitschrift für Pädagogik*, 61. Beiheft, 11–31.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (Hrsg.) (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster/New York/München/Berlin: Waxmann.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Überarbeitete Fassung des Antrags an die DFG auf Einrichtung eines Schwerpunktprogramms. *Zeitschrift für Pädagogik*, 52(6), 876–903.
- KMK Kultusministerkonferenz (2004). *Standards für die Lehrerbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz vom 16. 12. 2004*. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf [29. 05. 2014].
- KMK Kultusministerkonferenz (2005). *Qualifikationsrahmen für Deutsche Hochschulabschlüsse*. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_04_21-Qualifikationsrahmen-HS-Abschluesse.pdf [29. 05. 2014].
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster/New York/München/Berlin: Waxmann.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2001). *Intelligenz-Struktur-Test 2000R*. Göttingen: Hogrefe.
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch measurement transactions*, 7(4), 328.
- Liu, O. L. (2011). Value-added assessment in higher education: a comparison of two methods. *Higher Education*, 61(4), 445–461.
- McMillan, J. H., & Schumacher, S. (2010). *Research in education. Evidence-based inquiry* (7. Aufl.). Upper Saddle River: Pearson.
- Neisser, U., Boodoo, G., Bouchard, Th. J., Wade Boykin, A., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77–101 [DOI: 10.1037/0003-066X.51.2.77].

- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. Los Angeles: Cambridge Assessment/Sage.
- Prenzel, M., Walter, O., & Frey, A. (2007). PISA misst Kompetenzen. *Psychologische Rundschau*, 58(2), 128–136 [DOI: 10.1026/0033-3042.58.2.128].
- Ramseier, E., & Brühwiler, C. (2003). Herkunft, Leistung und Bildungschancen im gegliederten Bildungssystem. Vertiefte PISA-Analyse unter Einbezug der kognitiven Grundfähigkeiten. *Schweizerische Zeitschrift für Bildungswissenschaften*, 25(1), 23–58.
- Rauch, D., & Hartig, J. (2012). Interpretation von Testwerten in der IRT. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 253–264). Berlin: Springer.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 57(2), 69–86 [DOI: 10.1026/0033-3042.57.2.69].
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research – Online*, 8, 23–74.
- Schladitz, S., Rott, B., Winter, A., Wischgoll, A., Groß Ophoff, J., Hosenfeld, I., Leuders, T., Nückles, M., Renkl, A., Stahl, E., Watermann, R., Wirtz, M., & Wittwer, J. (2013). LeScEd – Learning the Science of Education. Research Competence in Educational Sciences. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *The German funding initiative „Modeling and Measuring Competencies in Higher Education“: 23 research projects on engineering, economics and social sciences, education and generic skills of higher education students* (KoKoHs Working Papers No. 3, S. 82–84). Berlin/Mainz: Humboldt-Universität/Johannes Gutenberg-Universität.
- Shank, G. D., & Brown, L. (2007). *Exploring educational research literacy*. New York: Routledge.
- Spinath, B., & Steinmayr, R. (2012). The roles of competence beliefs and goal orientations for change in intrinsic motivation. *Journal of Educational Psychology*, 104(4), 1135–1148 [DOI: 10.1037/a0028115].
- Süß, H.-M. (2003). Intelligenztheorien. In K. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 217–224). Weinheim: Psychologie Verlags Union.
- Taub, G. E., Floyd, R. G., Keith, T. Z., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly*, 23(2), 187–198 [DOI: 10.1037/1045-3830.23.2.187].
- Terhart, E. (2012). „Bildungswissenschaften“. Verlegenheitslösung, Sammelkategorie, Kampfbegriff? *Zeitschrift für Pädagogik*, 58(1), 22–39.
- Trapmann, S. (2008). *Mehrdimensionale Studienerfolgsprognose: Die Bedeutung kognitiver, temperamentsbedingter und motivationaler Prädiktoren für verschiedene Kriterien des Studienerfolgs*. Berlin: Logos.
- Urban, D., & Mayerl, J. (2008). *Regressionsanalyse: Theorie, Technik und Anwendung*. Berlin/Heidelberg: Springer.
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Hrsg.), *IERI monograph series: Issues and methodologies in large scale assessments, Vol. 2* (S. 9–36). Hamburg/Princeton: IERInstitute.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Seattle: Hogrefe & Huber.
- Weiß, R. H. (2006). *CFT 20-R. Grundintelligenztest – Revision*. Göttingen: Hogrefe.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy – value theory of motivation. *Contemporary Educational Psychology*, 25, 68–81 [DOI: 10.1006/ceps.1999.1015].
- Zlatkin-Troitschanskaia, O., & Kuhn, C. (2010). *Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen – Analyse zum Forschungsstand*. Mainz: Johannes Gutenberg-Universität.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly. An International Journal*, 4(2), 1–19.

Abstract: Being able to make evidence-based decisions is a central aim in higher education. The project LeScEd (Learning the Science of Education) aims at incorporating approaches from fields like Information Sciences, Mathematical Education and Psychology into a comprehensive structure model of Educational Research Literacy (ERL). The current study analyzes the relations of ERL to fluid intelligence and self-reported ERL to analyze discriminant and congruent validity of the developed instrument, respectively. Structural equation modelling revealed small positive effects regarding fluid intelligence and no effects regarding self-reported ERL. These results indicate that ERL is related to, but distinguishable from general intelligence and that self-reported competence is no reliable indicator for actual competence.

Keywords: Educational Research Literacy; Higher Education; Intelligence; Construct Validity; Structural Equation Modeling

Anschrift der Autorinnen/des Autors

Sandra Schladitz, Pädagogische Hochschule Freiburg, Institut für Psychologie,
Kunzenweg 21, 79117 Freiburg, Deutschland
E-Mail: sandra.schladitz@ph-freiburg.de

Dr. Jana Groß Ophoff, Pädagogische Hochschule Freiburg, Institut für Psychologie,
Kunzenweg 21, 79117 Freiburg, Deutschland
E-Mail: jana.grossophoff@ph-freiburg.de

Prof. Dr. Markus Wirtz, Pädagogische Hochschule Freiburg, Institut für Psychologie,
Kunzenweg 21, 79117 Freiburg, Deutschland
E-Mail: markus.wirtz@ph-freiburg.de